

**REMARKS**

In the non-final Office Action, dated August 22, 2007, the Examiner rejects claims 20-34 under 35 U.S.C. § 101 as allegedly being directed to non-statutory subject matter; rejects claims 1, 2, 5, 6, 8, 11, 13, 16, 17, 20, 24, 26, 27, and 28 under 35 U.S.C. § 102(a) as allegedly being anticipated by Schleimer et al., “Winnowing: Local Algorithms for Document Fingerprinting,” published June 9, 2003 (hereinafter “SCHLEIMER”); rejects claim 20 as allegedly being anticipated by U.S. Patent No. 5,745,900 B1 to Burrows (hereinafter “BURROWS”); rejects claims 1-2, 5-8, 13, 16-17, 24-29, 31, 32, and 34 under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS in view of U.S. Published Patent Application No. 2002/0133499 A1 to Ward et al. (hereinafter “WARD”); rejects claims 3, 4, 11, 12, 14 and 15 under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS and WARD and further in view of U.S. Patent No. 6,230,155 B1 to Broder et al. (hereinafter “BRODER”); rejects claims 22 and 23 under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS in view of Charikar, “Similarity Estimation Techniques from Rounding Algorithms”, published May 19, 2002 (hereinafter “CHARIKAR”); rejects claims 9, 18, and 33 under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS and WARD and further in view of Official Notice; and rejects claim 21 under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS in view of Official Notice. Applicant respectfully traverses these rejections.<sup>1</sup>

---

<sup>1</sup> As Applicant's remarks with respect to the Examiner's rejections are sufficient to overcome these rejections, Applicant's silence as to assertions by the Examiner in the Office Action or certain requirements that may be applicable to such rejections (e.g., whether a reference constitutes prior art, motivation to combine references, assertions as to dependent claims, etc.) is not a concession by Applicant that such assertions are accurate or such requirements have been met, and Applicant reserves the right to analyze and dispute such assertions/requirements in the future.

The Examiner objected to claim 19 as being dependent upon a rejected base claim, but indicated that claim 19 would be allowable if rewritten in independent form. Applicant notes with appreciation this indication of allowable subject matter.

By way of present amendment, Applicant amends claims 1-2, 5-11, 13 and 20-34 to improve form. No new matter has been added by way of the present amendment. Claims 1-34 remain pending.

Claims 20-34 stand rejected under 35 U.S.C. § 101 as being allegedly directed to non-statutory subject matter. Applicant respectfully traverses this rejection.

The Examiner alleged that the features of independent claims 20, 26, and 34 are intended to be implemented in software and that the claims lack the necessary physical articles or objects to constitute a machine or a manufacture as required by 35 U.S.C. § 101. Without necessarily agreeing with the Examiner, Applicant has amended claims 20-30 and 34 to recite “a computer-implemented device,” and amended claims 31-33 to recite a “computer-readable memory device” to address the Examiner’s concerns.

For at least the foregoing reasons, Applicant requests that the rejection of claims 20-34 under 35 U.S.C. § 101 be reconsidered and withdrawn.

Claims 1, 2, 5, 6, 8, 11, 13, 16, 17, 20, 24, 26, 27, and 28 stand rejected under 35 U.S.C. § 102(a) as allegedly being anticipated by SCHLEIMER. Applicant respectfully traverses this rejection.

Independent claim 1 is directed to a method for generating a representation of a document. The method includes sampling the document to obtain a plurality of overlapping blocks, choosing a subset of the plurality of overlapping blocks, and compacting the subset of the plurality of overlapping blocks to obtain the representation

of the document. SCHLEIMER does not disclose or suggest this combination of features.

For example, SCHLEIMER does not disclose or suggest choosing a subset of the plurality of overlapping blocks, as recited in claim 1. The Examiner relies on Section 3 and Fig. 2(e) of SCHLEIMER for allegedly disclosing this feature (Office Action, p. 4). Applicant disagrees with the Examiner's interpretation of SCHLEIMER.

Section 3 of SCHLEIMER describes the winnowing algorithm for selecting fingerprints from hashes of  $k$ -grams. In the example of Fig. 2(a)-2(g) of SCHLEIMER, a set of overlapping 5-grams is generated from a set of text (Fig. 2(c) of SCHLEIMER). A sequence of hashes is then generated from the sequence of 5-grams (Fig. 2(d) of SCHLEIMER). Next, a sequence of overlapping windows of the hashes is generated (Fig. 2(e) of SCHLEIMER), followed by the selection of the minimum hash from each window (Fig. 2(f) of SCHLEIMER). This sequence of minimum hashes becomes the fingerprint of the text. This section of SCHLEIMER does not disclose or suggest choosing a subset of a plurality of overlapping blocks that is obtained by sampling a document, as recited in claim 1. Instead, SCHLEIMER appears to hash all of the 5-grams derived from the text (see Figs. 2(c) and 2(d)), and does not choose a subset of the 5-grams.

The Examiner further relies on SCHLEIMER's alleged disclosure of "windows can be overlapping sample of a document" as allegedly corresponding to the feature of sampling the document to obtain a plurality of overlapping blocks, as recited in claim 1 (Office Action, p. 4). The Examiner is apparently referring to the windows of hashes shown in Fig. 2(e) of SCHLEIMER. The Examiner also appears to rely on Fig. 2(e) for

allegedly corresponding to choosing a subset of the plurality of overlapping blocks, as recited in amended claim 1 (Office Action, p. 4).

At the outset, Applicant submits that the windows of hashes depicted in Fig. 2(e) of SCHLEIMER are not obtained by sampling the document, as would be required by claim 1 based on the Examiner's interpretation of SCHLEIMER, but are rather obtained from sequences of hashes generated from the 5-grams.

Furthermore, it is unclear how the Examiner can rely on the windows of hashes shown in Fig. 2(e) of SCHLEIMER as allegedly corresponding to sampling the document to obtain the plurality of overlapping blocks recited in claim 1, and at the same time rely on the windows of the hashes in Fig. 2(e) of SCHLEIMER as also corresponding to choosing a subset of the plurality of overlapping blocks recited in claim 1. SCHLEIMER does not disclose or suggest sampling the document to obtain overlapping windows of hashes and choosing a subset of the overlapping windows of hashes, as would be required by claim 1, based on the Examiner's apparent interpretation of SCHLEIMER. SCHLEIMER only discloses choosing the minimum hash from each window, and does not disclose choosing a subset of the windows of hashes.

For at least the foregoing reasons, Applicant submits that claim 1 is not anticipated by SCHLEIMER.

Claims 2, 5, 6, 8, and 11 depend from claim 1. Therefore these claims are not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim 1. Moreover, these claims are not anticipated by SCHLEIMER for reasons of their own.

For example, claim 2 recites that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on

the subset of the plurality of overlapping blocks. SCHLEIMER does not disclose or suggest this feature.

The Examiner relies on Fig. 2(g) of SCHLEIMER as allegedly disclosing this feature (Office Action, p. 5). Applicant disagrees with the Examiner's interpretation of SCHLEIMER. Fig. 2(g) of SCHLEIMER discloses pairing the fingerprint generated by selecting the minimum hash from each hash window of Fig. 2(e) with a number indicating the position of this hash in the sequence of hashes generated from the 5-grams (as shown in Fig. 2(d) of SCHLEIMER). The Examiner relies on the fingerprint of Fig. 2(g) of SCHLEIMER as allegedly corresponding to representation of the document, as recited in claim 1. SCHLEIMER does not disclose or suggest setting the bits of the fingerprint based on the subset of the windows of hashes, as would be required by claim 2, based on the Examiner's interpretation of SCHLEIMER. In fact, SCHLEIMER does not disclose or suggest setting the bits of the fingerprint at all. SCHLEIMER only discloses pairing each hash of the fingerprint with a number indicating the position of the hash. Therefore, SCHLEIMER cannot disclose or suggest that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks, as recited in claim 2.

For at least these additional reasons, Applicant submits that claim 2 is not anticipated by SCHLEIMER.

Independent claim 13 is directed to a method for generating a representation of a document. The method includes sampling the document to obtain a plurality of overlapping samples, selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest

samples or a predetermined number of largest samples, and setting bits in the representation of the document based on the selected predetermined number of the samples. SCHLEIMER does not disclose or suggest this combination of features.

The Examiner did not address the features of claim 13 in the Office Action. Instead, the Examiner addressed the features of claim 1 (Office Action, p. 5). Claim 1 does not, for example, recite selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, as recited in claim 13. The Examiner does not address this feature. Thus, a proper case of anticipation has not been established with respect to claim 13. If this rejection is maintained, Applicant respectfully requests that the Examiner specifically address the features recited in claim 13.

Applicant further submits that SCHLEIMER does not, for example, disclose or suggest setting bits in the representation of the document based on the selected predetermined number of the samples, as recited in claim 13. This feature is similar to the feature present in claim 2, which was addressed above.

For at least these reasons, Applicant submits that claim 13 is not anticipated by SCHLEIMER.

Claims 16 and 17 depend from claim 13. Therefore, these claims are not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim 13.

Amended independent claims 20 and 26 recite features which are similar to, yet possibly of different scope than, the features recited above with respect to claim 1.

Therefore, these claims are not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim 1.

Claim 24 depends from claim 20. Therefore, this claim is not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim 20.

Claims 27 and 28 depend from claim 26. Therefore, these claims are not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim 26.

Claim 20 stands rejected as being allegedly anticipated by BURROWS. Applicant respectfully traverses this rejection.

Amended independent claim 20 is directed to a computer-implemented device. The device includes a fingerprint creation component to generate a fingerprint of a predetermined length for an input document, the fingerprint generated by sampling the input document to obtain samples, choosing a subset of the samples, and generating the fingerprint from the subset of the samples by compacting the subset of the samples, and a similarity detection component to compare pairs of fingerprints to determine whether the pairs of fingerprints correspond to near-duplicate documents. BURROWS does not disclose or suggest this combination of features.

For example, BURROWS does not disclose or suggest a fingerprint creation component to generate a fingerprint of a predetermined length for an input document, the fingerprint generated by sampling the input document to obtain samples, choosing a subset of the samples, and generating the fingerprint from the subset of the samples by compacting the subset of the samples, as recited in claim 20. The Examiner relies on Figs. 4 and 5 of BURROWS for allegedly disclosing this feature (Office Action, p. 7).

Applicant disagrees with the Examiner's interpretation of BURROWS.

Fig. 4 of BURROWS, which is described in col. 7, line 41 to col. 9, line 32, depicts a block diagram of content attributes generated by the search engine. Fig. 4 shows portions of a page, labeled as 230, 240, 250, and 260, which are detected and encoded by a parsing module. The parsing module generates attribute values for entire pages, portions of a page, fields, or individual words and the parser stores these attribute values as searchable metawords. The fingerprint 255 of Fig. 4 can be one of these metawords. BURROWS specifically discloses that fingerprint 255 can be produced by applying one-way polynomial functions to the digitized content of the document (col. 8, lines 16-23 of BURROWS). BURROWS in no way discloses or suggests a fingerprint creation component to generate a fingerprint of a predetermined length for an input document, the fingerprint generated by sampling the input document to obtain samples, choosing a subset of the samples, and generating the fingerprint from the subset of the samples by compacting the subset of the samples, as recited in claim 20.

Fig. 5 of BURROWS, described in col. 9, lines 33-41, shows a view of the words and metawords produced by the parsing module. The parsing module produces a sequence of pairs in a collating order according to the location of the words of various pages.

Fig. 5 of BURROWS does not disclose or suggest anything about a fingerprint creation component. The only section of BURROWS that discloses anything about a fingerprint is col. 8, lines 16-24, which describes fingerprint 255 consisting of an integer value generated by applying a one-way polynomial, as described above. Therefore, this figure of BURROWS cannot disclose or suggest a fingerprint creation component to



generate a fingerprint of a predetermined length for an input document, the fingerprint generated by sampling the input document to obtain samples, choosing a subset of the samples, and generating the fingerprint from the subset of the samples by compacting the subset of the samples, as recited in claim 20.

For at least the foregoing reasons, Applicant submits that claim 20 is not anticipated by BURROWS.

Claims 1-2, 5-8, 13, 16-17, 24-29, 31-32, and 34 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over BURROWS in view of WARD. Applicant respectfully traverses this rejection.

Independent claim 1 is directed to a method for generating a representation of a document. The method includes sampling the document to obtain a plurality of overlapping blocks, choosing a subset of the plurality of overlapping blocks, and compacting the subset of the plurality of overlapping blocks to obtain the representation of the document. BURROWS and WARD, whether taken alone or in any reasonable combination, do not disclose or suggest this combination of features.

For example, BURROWS and WARD do not disclose or suggest choosing a subset of the plurality of overlapping blocks, as recited in claim 1. The Examiner relies on Fig. 4 of BURROWS for allegedly disclosing this feature (Office Action, p. 8).

Fig. 4 of BURROWS, which is described in col. 7, line 41 to col. 9, line 32, depicts a block diagram of content attributes generated by the search engine. Fig. 4 shows portions of a page, labeled as 230, 240, 250, and 260, which are detected and encoded by a parsing module. The parsing module generates attribute values for entire pages, portions of a page, fields, or individual words and the parser stores these attribute

values as searchable metawords. Fingerprint 255 of Fig. 4 can be one of these metawords. BURROWS specifically discloses that the fingerprint 255 can be produced by applying one-way polynomial functions to the digitized contents of the document (col. 8, lines 16-23 of BURROWS).

The Examiner apparently relies on the portions of the pages 230, 240, 250, and 260 of Fig. 4 as allegedly corresponding to the plurality of overlapping blocks, as recited in claim 1, and relies on Fig. 4 as allegedly corresponding to choosing a subset of the plurality of overlapping blocks, as recited in claim 1. However, Fig. 4 of BURROWS does not disclose or suggest choosing a subset of the portions of the pages 230, 240, 250, and 260, as would be required by claim 1, based on the Examiner's interpretation of BURROWS. Thus, BURROWS in no way discloses choosing a subset of a plurality of overlapping blocks obtained by sampling a document, as recited in claim 1.

Furthermore, BURROWS and WARD do not disclose or suggest compacting the subset of the plurality of overlapping blocks to obtain the representation of the document, as also recited in claim 1. The Examiner relies on Fig. 5 of BURROWS for allegedly disclosing this feature (Office Action, p. 8). The Examiner disagrees with the Examiner interpretation of BURROWS.

Fig. 5 of BURROWS, described in col. 9, lines 33-41, shows a view of the words and metawords produced by the parsing module. The parsing module produces a sequence of pairs in a collating order according to the location of the words of various pages. BURROWS does not disclose or suggest anywhere that this sequence of pairs in a collating order is a representation of the document. Therefore, this section of BURROWS does not disclose or suggest compacting the subset of the plurality of

overlapping blocks to obtain the representation of the document, as recited in claim 1.

WARD does not overcome the deficiencies of BURROWS set forth above with respect to claim 1. For example, WARD does not disclose or suggest choosing a subset of the plurality of overlapping blocks, and compacting the subset of the plurality of overlapping blocks to obtain the representation of the document, as recited in claim 1.

With regard to motivation, the Examiner alleges (Office Action, p. 8):

It would have been obvious to a person of ordinary skill in the art to use the window overlapping sampling of Ward et al. with the method of Burrows because it allows for quicker indexing a higher accuracy of the resulting samples.

Applicant submits that the Examiner's allegation is merely a conclusory statement regarding an alleged benefit of the combination of BURROWS and WARD. Such conclusory statements are insufficient for establishing a *prima facie* case of obviousness.

For at least the foregoing reasons, Applicant submits that claim 1 is patentable over BURROWS and WARD, whether taken alone or in any reasonable combination.

Claims 2 and 5-8 depend from claim 1. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 1. Moreover, these claims are patentable over BURROWS and WARD for reasons of their own.

For example, claim 2 recites that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks. BURROWS and WARD do not disclose or suggest this feature.

The Examiner relies on Fig. 5 of BURROWS for allegedly disclosing this feature (Office Action, p. 8). Fig. 5 of BURROWS, described in col. 9, lines 33-41, shows a view of the words and metawords produced by the parsing module. The parsing module

produces a sequence of pairs in a collating order according to the location of the words.

Neither this figure nor the description thereof discloses or suggests that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks, as recited in claim 2.

WARD does not overcome the deficiencies of BURROWS set forth above with respect to claim 2.

For at least these additional reasons, Applicant submits that claim 2 is patentable over BURROWS and WARD, whether taken alone or in any reasonable combination.

Independent claim 13 is directed to a method for generating a representation of a document. The method includes sampling the document to obtain a plurality of overlapping samples, selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, and setting bits in the representation of the document based on the selected predetermined number of the samples. BURROWS and WARD do not disclose or suggest this combination of features.

The Examiner did not address the features of claim 13 in the Office Action. Instead, the Examiner addressed the features of claim 1 (Office Action, p. 11). Claim 1 does not, for example, recite selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, as recited in claim 13. The Examiner does not address this feature. Thus, a *prima facie* case of obviousness has

not been established with respect to claim 13. If this rejection is maintained, Applicant respectfully requests that the Examiner specifically address the features recited in claim 13.

Nevertheless, Applicant submits that BURROWS and WARD do not disclose or suggest selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, as recited in claim 13.

For at least the foregoing reasons, Applicant submits that claim 13 is patentable over BURROWS and WARD, whether taken alone or in any reasonable combination.

Claims 16-17 depend from claim 13. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 13.

Claims 24 and 25 depend from claim 20. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 20.

Independent claims 26 and 34 recite features which are similar to, yet possibly of different scope than, the features recited above with respect to claim 1. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 1.

Claims 27-29 depend from claim 26. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 26.

Independent claim 31 recites features which are similar to, yet possibly of different scope than, the features recited above with respect to claim 13. Therefore, this claim is patentable over BURROWS and WARD for at least the reasons set forth above

with respect to claim 13.

Claim 32 depends from claim 31. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 31.

Claims 3, 4, 11, 12, 14 and 15 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over BURROWS and WARD and further in view of BRODER. Applicant respectfully traverses this rejection.

Claims 3, 4, 11, and 12 depend from claim 1. BRODER does not overcome the deficiencies of BURROWS and WARD set forth above with respect to claim 1. Therefore, these claims are patentable over BURROWS, WARD, and BRODER, for at least the reasons set forth above with respect to claim 1.

Claims 14 and 15 depend from claim 13. BRODER does not overcome the deficiencies of BURROWS and WARD set forth above with respect to claim 13. Therefore, these claims are patentable over BURROWS, WARD, and BRODER, for at least the reasons set forth above with respect to claim 13.

Claims 22 and 23 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS in view of CHARIKAR. Applicant respectfully traverses this rejection.

Claims 22 and 23 depend from claim 20. CHARIKAR does not overcome the deficiencies of BURROWS set forth above with respect to claim 20. Therefore, these claims are patentable over BURROWS, and CHARIKAR, for at least the reasons set forth above with respect to claim 20.

Claims 9, 18, and 33 stand rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS and WARD in view of Official Notice. Applicant

respectfully traverses this rejection.

Claim 9 depends from claim 1. The Examiner's taking of Official Notice does not overcome the deficiencies of BURROWS and WARD set forth above with respect to claim 1. Therefore, this claim is patentable over BURROWS, WARD and Official Notice for at least the reasons set forth above with respect to claim 1.

Claim 18 depends from claim 13. The Examiner's taking of Official Notice does not overcome the deficiencies of BURROWS and WARD set forth above with respect to claim 13. Therefore, this claim is patentable over BURROWS, WARD and Official Notice for at least the reasons set forth above with respect to claim 13.

Claim 33 depends from claim 31. The Examiner's taking of Official Notice does not overcome the deficiencies of BURROWS and WARD set forth above with respect to claim 31. Therefore, this claim is patentable over BURROWS, WARD and Official Notice for at least the reasons set forth above with respect to claim 31.

Claim 21 stands rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS and in view of Official Notice. Applicant respectfully traverses this rejection.

Claim 21 depends from claim 20. The Examiner's taking of Official Notice does not overcome the deficiencies of BURROWS set forth above with respect to claim 20. Therefore, this claim is patentable over BURROWS and Official Notice for at least the reasons set forth above with respect to claim 20.

In view of the foregoing amendments and remarks, Applicant respectfully requests the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

While the present application is believed to be in condition for allowance, should the Examiner find some issue to remain unresolved, or should any new issues arise that could be eliminated through discussions with Applicant's representative, then the Examiner is invited to contact the undersigned by telephone to expedite prosecution of the present application.

To the extent necessary, a petition for an extension of time under 37 CFR § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

By: /Viktor Simkovic, Reg. No. 56012/  
Viktor Simkovic  
Registration No. 56,012

Date: November 21, 2007

Harrity Snyder, L.L.P.  
11350 Random Hills Road  
Suite 600  
Fairfax, Virginia 22030  
Main: (571) 432-0800  
Direct: (571) 432-0899  
Customer Number: 44989